

MENU

SEARCH

INDEX

DETAIL

JAPANESE

BACK

NEXT

5 / 18

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-244521
 (43)Date of publication of application : 08.09.2000

(51)Int.Cl. H04L 12/28
 H04L 12/56

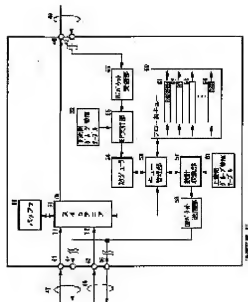
(21)Application number : 11-042975 (71)Applicant : SUMITOMO ELECTRIC IND LTD
 (22)Date of filing : 22.02.1999 (72)Inventor : MURATA TAKUSHI

(54) METHOD AND DEVICE FOR COMMUNICATION

(57)Abstract:

PROBLEM TO BE SOLVED: To efficiently report flow control based on a back pressure(BP) while collecting plural flows into group.

SOLUTION: Packets inputted to input ports 41 and 42 of communication equipment 35 are stored in a buffer 52 and a packet descriptor is stored in queues 61-64 for each flow. The group collecting plural flows is shown in an upstream side group reference table. On the basis of the packet storage quantity of the group, etc., a statistic collecting part 57 sends a BP packet through a BP packet transmitting part 58 to upstream side communication equipment. On the other hand, the BP packet sent from downstream side communication equipment is received by a BP packet receiving part 55. The flows included in the group are shown in a downstream side group reference table 82 and the control of flows is performed by a BP executing part 56.



LEGAL STATUS

[Date of request for examination] 20.10.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3107077

[Date of registration] 08.09.2000

[Number of appeal against examiner's decision of
rejection]

[Date of requesting appeal against examiner's decision of
rejection]

[Date of extinction of right]

08.09.2004

1

【特許請求の範囲】

【請求項1】 複数のフローを含むリンクによって互いに接続された通信装置間でフローを制御する通信方法において、

予め、複数のフローをまとめたグループを定め、通信装置間で前記グループを通知しておき、下流の通信装置から上流の通信装置にフローの停止又は停止解除を通知するときは、前記グループによって通知することで、そのグループに含まれるフローに対して一括してフローの停止又は停止解除を通知することを特徴とする通信方法。

【請求項2】 複数のフローを含むリンクによって互いに接続される通信装置において、複数のフローをまとめたグループを示す下流側グループ参照手段と、下流の通信装置からグループによってフローの停止又は停止解除の通知を受けるフロー制御受信手段と、前記グループに含まれるフローに対してフローの停止又は停止解除を行うフロー制御実行手段とを有することを特徴とする通信装置。

【請求項3】 複数のフローを含むリンクによって互いに接続され、フローを構成するパケット又はセル（以下、パケットで代表する）を入力ポートから入力し一旦記憶領域に蓄えた後に出力ポートから出力することで通信を行う通信装置において、複数のフローをまとめたグループを示す上流側グループ参照手段と、グループに含まれるフローを構成するパケットに関して前記記憶領域に蓄えられているパケット記憶量を求める統計手段と、

前記パケット記憶量が停止閾値より多くなったときは上流の通信装置に前記グループによってフローの停止を通知し、前記パケット記憶量が再開閾値より少なくなったときは上流の通信装置に前記グループによってフローの停止解除を通知するフロー制御送信手段とを有することを特徴とする通信装置。

【請求項4】 複数のフローを含むリンクによって互いに接続され、フローを構成するパケット又はセル（以下、パケットで代表する）を入力ポートから入力し一旦記憶領域に蓄えた後に出力ポートから出力することで通信を行う通信装置において、複数のフローをまとめたグループを示す上流側グループ参照手段であって、パケットの廃棄よりも遅配が望ましいグループ（以下、遅配グループ）と、パケットの遅配よりも廃棄が望ましいグループ（以下、廃棄グループ）とを示すものと、

グループに含まれるフローを構成するパケットに関して前記記憶領域に蓄えられているパケット記憶量を求める統計手段であって、前記遅配グループと前記廃棄グループそれぞれに関してパケット記憶量を求めるものと、前記遅配グループのパケット記憶量が停止閾値より多く

2

なったときは上流の通信装置に前記グループによってフローの停止を通知し、前記パケット記憶量が再開閾値より少なくなったときは上流の通信装置に前記グループによってフローの停止解除を通知するフロー制御送信手段と、

前記廃棄グループのパケット記憶量が廃棄閾値より多くなったときは廃棄グループに含まれるフローのパケットを廃棄するパケット廃棄手段とを有することを特徴とする通信装置。

【請求項5】 廃棄されやすさの異なるパケットが前記廃棄グループのフローに含まれており、前記パケット廃棄手段は、パケットに含まれている情報を基に予め定められたきまりに従って廃棄されやすさを決定し、前記パケット記憶量に応じて廃棄されやすいパケットを優先的に廃棄することを特徴とする請求項4に記載の通信装置。

【請求項6】 前記廃棄閾値が、前記記憶領域の空き容量に応じて変化することを特徴とする請求項4又は5に記載の通信装置。

【請求項7】 同一の出力ポートを通る複数のフローをまとめたグループがあり、各グループには、パケット記憶量の上限と、前記パケット記憶量の上限よりも小さい停止閾値及び再開閾値を備えることを特徴とする請求項3から6に記載の通信装置。

【請求項8】 グループに対するパケット記憶量が停止閾値より多くなった後、上流の通信装置にフローの停止を通知して、実際にフローが停止するまでの遅延時間を停止遅延時間とし、

前記上流の通信装置が接続されている入力ポートから流入されるパケットの単位時間当たりの伝送量を伝送帯域とし、

入力ポート毎に決まる停止遅延時間と伝送帯域の積に相当する容量を停止遅延相当分容量とし、

グループに含まれるフローに関連する全入力ポートの停止遅延相当分容量の和に相当する容量をグループ停止遅延相当容量以上であることを特徴とする請求項7に記載の通信装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、デジタルデータを伝送する際のパケットやA/T M (Asynchronous Transfer Mode：非同同期伝送モード)セルを中継する通信装置に関する。

【0002】

【従来の技術】 コンピュータやデータ端末など（以下、コンピュータで代表する）を通信装置で接続してネットワークを形成し、各コンピュータ間でデータを伝送する

50

3
ことが行われている。図4にその構成を例示する。コンピュータ30〜34は、直接的には通信装置35〜38に接続されている。通信は双方向に行われるので、接続には少なくとも一対の伝送路(リンク)が必要である。接続は、複数の伝送路を用いても良く、ネットワーク全体として迂回路やループがあっても良い。図4には示していないが、電気信号を増幅するリピータなど他の種類の通信装置が、伝送路の途中に存在していても良い。図5と対比して説明する場合上、図4に示す通信装置では、左側に入力ポートを、右側出力ポートを記載している。通信装置35〜38は、互いに接続されており、コンピュータからの通信を他のコンピュータ又は通信装置に伝送する。この様に、図4に示すネットワークでは、通信装置を介してコンピュータとコンピュータの間で通信が行われる。

【0003】この様に通信装置で中継することにより、コンピュータ間での通信が行われる。通信は、データをパケットに区切り、パケットを単位として伝送することで行う。パケットは可変長の場合や固定長の場合もある。ATM技術を使う場合は、固定長のセルを単位として伝送する。本発明に関する通信装置は、パケットを伝送する場合でもセルを伝送する場合でも適用可能である。従って、以降では、セルの場合を含めて、パケットで代表して説明する。

【0004】この様に、通信が行われると、コンピュータとコンピュータの間にパケット列が流れる。同様に、コンピュータと通信装置の間、通信装置と通信装置の間、あるいは、コンピュータ上のアプリケーションとコンピュータ上のアプリケーションの間にパケット列が流れると言うこともできる。この様な特定の送信元と送信先の組の間に流れるパケット列をフローと呼ぶ。言うまでもなく、一つのリンクには複数のフローが流れる。リンクでは、予め定められたきまりに従ってリンクを流れるパケットを分類することによってフローを特定できる。分類はパケットに含まれる情報に基づいて行われる。分類のきまりの例には、パケット中の送信元と送信先の物理アドレスの組でフローを特定するというきまりや、ATMセル中の仮想バス識別子と仮想チャネル識別子の組でフローを特定するといったきまりなどを挙げる事ができる。

【0005】通信の対象となるデータには、リアルタイムの動画や音声の様に、通信の遅延や伝送レートが一定以下に劣化するとか固めるものがある。また、一方、ファイル転送の様ななるべく高い伝送レートで送りたいが、遅延や伝送レートの劣化があっても差し支えないものがある。アプリケーションによって求められるQoS(Quality of Service: サービス品質)が異なる訳である。例えば、前者は、伝送レートが一定以上に、遅延や遅延の揺らぎが一定以下に保たれる必要がある。後者は、伝送レートや遅延が一定に保たれる必要がない。ただし、リ

4
ンクの未使用帯域があれば、その未使用帯域を使用させることによって、できるだけ高い伝送レート(伝送帯域)で通信させる方が好ましい。すなわち、ネットワークの使用効率をなるべく高めたい。

【0006】そこで、通信装置においては、一定のQoSを保証すべき通信量に対しては、通信装置の伝送帯域や一時記憶領域といった資源を予め割り当てる。資源を割り当てた後、割り当てた資源が実際に使われていない場合や、そもそも余剰の資源がある場合には、未使用分の資源を一定のQoSを保証すべき通信量以外に相当する部分の通信に資源を割り当てるとネットワークの仕様効率を高めることができる。

【0007】前記の資源が伝送帯域である場合の資源を割り当て方を例示する。例えば、前者の通信(通信の遅延や伝送レートが一定以下に劣化するとか固めるもの)には、最高帯域(ATM技術の場合はPeak Cell Rateに相当)を保証するように伝送帯域を割り当てる。一方、後者の通信(なるべく高い伝送レートで送りたいが、遅延や伝送レートの劣化があっても差し支えないもの)には最低帯域(ATM技術の場合はMinimum Cell Rateに相当)を保証するように伝送帯域を割り当てる。なお、最低帯域として、零の最低帯域を保証しても良い。この様に、通信の対象となるデータの性質に合わせて伝送帯域を割り当てることで、ネットワークの使用効率を高めることができる。

【0008】前記未使用分の伝送帯域をできるだけ多く使って通信を提供するようサービスを提供する。以後ベスト・エフォート・サービスと呼ぶ。なお、本発明に関する通信装置は、最低帯域の保証を寧ろ設定しない。また、本発明に関する通信装置は、ベスト・エフォート・サービスを適用しないフロー(例えば、常に一定の伝送レートを保つ帯域予約型フロー)を含んでいても良い。言うまでもなく、帯域予約型フローはベスト・エフォート・サービスを適用するフローの影響を受けない。

【0009】さて、資源が予め割り当てられていないフローにおいて、途中でパケットが廃棄されることなく装置間で通信が行われるためには、パケットが流れる下流の装置が受け取ることができる量に合わせて上流の装置が送信を行わなければならない。これをフロー制御と呼ぶ。具体的には、下流の装置がパケットを受け取るか否かといった輻輳状態に関する情報をその上流の装置にフィードバックすることで行う。例えば、下流の装置内に設けてあるパケットを受け取るバッファ(一時的な記憶領域)に溜まっているパケットの量を監視し、所定の閾値(停止閾値)より多くなれば、下流の装置から上流の装置へ送信を停止するよう通知する。逆に、バッファに溜まっているパケットの量が、別の所定の閾値(再開閾値)より少なくなれば、下流の装置から上流の装置へ送信を停止解除(再開)するよう通知する。この様な、下流の状態による上流の送信の制御をバックプレッシャ

5

と呼ぶ。バックプレッシャによるフロー制御は、例えば、特開平8-214004号に開示されている。

【0010】下流の状態による上流の送信の制御をおおまかに分類すると、隣接する装置間で行うlink-by-linkと、途中の経路を含めて最終的なパケットの送受信装置間で行うend-to-endの2通りある。本発明に関する通信装置ではlink-by-linkのバックプレッシャを行う。以下では、バックプレッシャによる送信の停止・停止解除を通知するパケットを、バックプレッシャパケットと呼ぶ。バックプレッシャに基づくフロー制御を行うためには、上流の装置から下流の装置にパケットが送信されている最中に下流の装置から上流の装置に輻輳状態に関する情報をフィードバックしなければならない。例えば、フィードバックはバックプレッシャパケットを送付することや、上流へ向かうパケットのどこかに埋め込ませることによって行うことができる。これら上流の装置と下流の装置をつなぐリンクは、全二重(full duplex)モードで動作させることが一般的である。

【0011】ところで、通信装置の間をつなぐ伝送路には、コンピュータ間の通信が複数含まれる。フローに必要なQoSは前述した様に種々に異なる。すなわち、QoSの異なる複数のフローが同じ通信装置の経路を通る。また、バックプレッシャの生じ方は個々のフロー毎に異なる。従って、個々のフロー毎にバックプレッシャを制御できる必要がある。これを選択的バックプレッシャと呼ぶ。

【0012】従来技術により選択的バックプレッシャを行う通信装置35の構成を図5に示す。通信装置35は、n本の入力ポート41~43とn本の出力ポート44~46を備えており、n個の別の装置と通信を行うことができる。

【0013】51はスイッチコアであり、少なくともn+1本の入力ポート71~74と少なくともn+1本の出力ポート75~78を備える。n+1番目の入力ポート74(入力ポートx)と出力ポート78(出力ポートz)は、この通信装置35自身が、別の装置と通信するために用いる。例えば、選択的バックプレッシャに基づくフロー制御のためには、この通信装置と別の装置との連携が必要である。この入力ポートxと出力ポートzを使ってバックプレッシャパケットを通信することで、この連携を取る。

【0014】スイッチコア51に入力されたパケットは、どのフローに対応するものかが特定され、そのフローに対応する出力ポートから出力される。一つの出力ポートからは複数のフローが流出するが、一時点では一つのパケットしか出力できない。そこで、パケットを一旦フロー毎に用意した待ち行列(キュー)に蓄えておき、逐次的に、すなわち、時分割多重によって出力する。

【0015】パケットそのものをキューに蓄えることも

6

できる。しかし、この通信装置では、パケットに対してパケット記述子を付与し、このパケット記述子をキュー管理用メモリ60内に用意した各フロー毎のキュー61~64に蓄えて時分割処理を行っている。パケットは、全フローに共用する記憶領域であるバッファ52に記憶される。パケット記述子は、この通信装置内で生成し利用するものであり、バッファ52のどこにパケットを記憶したかを示している。

【0016】パケットに対する操作を詳述する。スイッチコア51に入力されたパケットは、どのフローに対応するものかが特定される。特定されたフローは、この通信装置内ではフロー識別子を付して表すことにする。パケットはバッファ52に記憶され、フロー識別子とパケット記述子がスイッチコア51からキュー管理部53に通知される。キュー管理部53は、スイッチコア51から受け取ったフロー識別子で示されるキュー(61~64のいずれかに)、パケット記述子を入れる(キューにパケット記述子を入れることを、以下では、キューイングとも呼ぶ)。このとき、パケット記述子が空のキューに対してキューイングされたのであれば、キュー管理部53は、フローが活性化したことをスケジューラ54に通知する。

【0017】キュー管理部53は、スケジューラ54からフロー識別子を受け取るとそのフローに対応するキューの先頭からパケット記述子を取り出す。このとき、キューが空になれば、フローが非活性化したことをスケジューラ54に通知する。また、キュー管理部53は、取り出したパケット記述子とそのキューに対応するフロー識別子をスイッチコア51に通知する。スイッチコア51は、そのパケット記述子に基づいてパケットをバッファ52から取り出し、フロー識別子に対応する出力ポートからそのパケットを出力する。以上の、パケット記述子の取り出しからパケットの出力までが逐次的に行われることで、パケットの時分割多重が行われている。

【0018】スケジューラ54の機能について、より詳しく説明する。スケジューラ54は、フロー識別子と出力ポートとの対応、出力ポートの所有する伝送帯域から各フローにどれだけの伝送帯域を割り当てるか、を管理する機能を有している。キューの中にパケット記述子がある場合(キューが空でない場合)は、フローが活性化されている場合であり、そのフローはスケジューラの処理(パケットを出力するタイミングを指示するサービス)の対象である。逆にキューが空の場合は、フローが非活性化されている場合であり、そのフローはスケジューラのサービスの対象でない。

【0019】スケジューラ54は、キュー管理部53から活性化されたフロー識別子を通知されると、そのフローをスケジューラ54のサービスの対象に加える。後述のバックプレッシャによるフローの停止を受けておらず、サービスの対象となっているフローに関しては、伝

50

7

送帯域を割り当てる所定の方針（ポリシー）に従って、パケットを取り出すタイミングをスケジューラ54は決める。このタイミングになれば、対応するフロー識別子をスケジューラ54はキュー管理部53に伝える。これによりキュー管理部53によるパケット記述子の取り出しやスイッチコア51によるパケットの出力が許可される。この際、もし、キュー管理部53から非活性化されたフローの通知を受けたなら、このフローはサービスの対象から除外される。

【0020】出力ポートの有する伝送帯域を、その出力ポートを通る各フローに割り当てるポリシーの代表的なものとして、パケットの場合はDDR (Deficit Round Robin) が、セルの場合はWRR (Weighted Round Robin) がある。これらは、各フローに重みを設定し、その重みの割合に従って、出力ポートの伝送帯域をフローの間で分け合うというものである。ラウンドロビンによる伝送帯域の割り当ては、例えば、サービスの対象となっている全てのフローを循環的に検査し、それぞれのフローからはフローの重みに応じてパケットを出力させることで実現する。最低保証帯域が正のフローがある場合は、一定時間毎に最低保証帯域を満たすようにパケットを出力し、残りの時間について、このラウンドロビンによる伝送帯域の割り当てを行ったり、各フローの重みを調節して最低保証帯域を満たすようにすれば良い。

【0021】下流からのバックプレッシャによるフロー制御は、B P (バックプレッシャ) パケット受信部55、B P (バックプレッシャ) 実行部56、スケジューラ54によって行われる。まず、下流からのバックプレッシャパケットはスイッチコアの出力ポートz78からB Pパケット受信部55へ渡される。B P実行部56は、B Pパケット受信部からバックプレッシャパケットの内容を受け取り、スケジューラ54に対してフローの停止、停止解除を指示する。スケジューラ54は、停止指示されたフローに対しては、以降のパケットの出力をキュー管理部53に指示しない。逆に、停止解除を指示されたフローに対しては、再びパケットの出力をキュー管理部53に指示するようになる。

【0022】上流へのバックプレッシャによるフロー制御は、キュー管理部53、統計収集部57、B Pパケット送信部58によって行われる。まず、キュー管理部53は、キュー61〜64の状態を統計収集部57に通知する。例えば、各キューに入っているパケット記述子をもとにキューイングされているパケットのデータ量（キュー長）をそのまま通知する。ATMの場合は、キュー長としてはセル数を通知すれば良い。可変長のパケットの場合は、パケットの数ではなくキューイングされているパケットのバケット長と和をキュー長として通知する。もし、可変長のパケットを切りの良い長さで切り上げて（量子化して）バッファ52に記憶するなら、量子化したパケット長の和をキュー長として通知する。ある

8

いは、キュー長をそのまま通知するのではなく、パケット記述子の個々の出し入れ単位にバケット長を統計収集部57に通知し、各キューに蓄えられているデータ量を統計収集部57の中において増減することで、キュー長を把握するのも良い。また、統計収集部57は、各キュー61〜64に蓄えられているパケットが占めるデータ量の総和を求めることで、バッファ52の使用量を知ることができる。

【0023】選択的バックプレッシャ、すなわち、個々のフロー毎のフロー制御は次の様に行う。統計収集部57は、各キュー61〜64のキュー長が停止閾値より多くなれば、対応するフローのパケットの出力を停止すべきと判断し、B Pパケット送信部58に通知する。B Pパケット送信部58は、そのフローの停止を要求するパケットを作成し、スイッチコア51を経てそのフローに対応する上流側の装置に送付する。また、統計収集部57は、キュー記憶量が再開閾値より小さくなれば、対応するフローのパケットの出力の停止を解除すべきと判断し、B Pパケット送信部58に通知する。B Pパケット送信部58は、そのフローの停止解除を要求するパケットを作成し、スイッチコア51を経てそのフローに対応する上流側の装置に送付する。

【0024】リンク全体のバックプレッシャ（グローバルバックプレッシャ）、すなわち、全フローに対する停止、停止解除の制御は次の様に行う。統計収集部57は、バッファ52に蓄えられているパケットの総量（パケット記憶量）が停止閾値より多くなれば、すべてのフローのパケットの出力を停止すべきと判断し、B Pパケット送信部58に通知する。B Pパケット送信部58は、全フローの停止を要求するパケットを作成し、スイッチコア51を経てすべての上流側の装置に送付する。また、統計収集部57は、パケット記憶量が再開閾値より小さくなれば、すべてのフローのパケットの出力の停止を解除すべきと判断し、B Pパケット送信部58に通知する。B Pパケット送信部58は、全フローの停止解除を要求するパケットを作成し、スイッチコア51を経てすべての上流側の装置に送付する。グローバルバックプレッシャパケットを受信した上流側装置では、スケジューラ54を停止させることによって上流側装置からのパケットの送信を停止させる。

【0025】言うまでもないが、選択的バックプレッシャによるフローの停止、又は、グローバルバックプレッシャによるフローの停止があれば、フローは停止させられる。逆に、選択的バックプレッシャによるフローの停止解除、及び、グローバルバックプレッシャによるフローの停止解除があれば、フローは再開させられる。

【0026】

【発明が解決しようとする課題】従来技術によるフローの制御は、個々のフローを対象とする選択的バックプレッシャと、全フローを対象とするグローバルバックプレ

ツシャの二種類しかない。

【0027】ところが、フローは幾つかの部分集合に分類することができる。例えば、一つの通信装置には複数の出力ポートがあるので、フローを出力ポートで分類する、すなわち、下流側の装置で分類することができる。あるいは、複数の送信先へ配信されるマルチキャストであるか、送信先が一つであるユニキャストであるかによってフローを分類することもできる。また、リアルタイムの動画や音声、と、ファイル転送では望ましいQoSが異なる様に、QoSによってフローを分類することができる。このように、フローは共通の属性によって、部分集合に分類することができ、望ましいフロー制御も異なっている。

【0028】ところが、かかるように特定の属性を共通に有する複数のフローに対して、フロー制御を一括して行うことは従来技術ではできなかった。もし仮に、個別のパックプレッシャを用いて、特定の属性を共通に有する複数のフローのフロー制御を行うならば、それぞれのフローを個々に指定しなければならぬ。すると、それぞれのフローに対してフロー制御用パケットを生成して送信しなければならぬ限り、フローの停止及び停止解除にフローの数だけ時間や伝送帯域を要してしまう。そして、パッファが満杯になる寸前に短時間にフローを停止させることができないので、それだけ余分にパッファの容量も必要になる。

【0029】本発明は、かかる場合のフロー制御を改善するために、全フロー中のある部分集合に相当するフローに対して一括してフロー制御を行う手段を提供することを第1の目的としている。

【0030】また、従来技術では、望ましいQoSに对应してフローを制御することが考慮されていない部分がある。例えば、リアルタイムの動画や音声ではパケットを遅配させるよりも廃棄することが望ましいことがあるのに対して、ファイル転送ではパケットを廃棄させるよりも遅配させることが望ましい。従来技術において複数のフローを一括して制御するグローバルバックプレッシャは、このようなQoSの相違を考慮せずに、すべてのフローを一括的に制御する。そのため、全フローを停止させると、パケットを遅配させるよりも廃棄することが望ましいフローも停止させられてしまう。すなわち、パケットの廃棄ではなく遅配が生じており、望ましいフロー制御にはなっていない。

【0031】本発明は、かかる場合のために、フローの属性に応じて一括してフローを制御できる手段を提供することを第2の目的としている。

【0032】また、従来技術では、一部の複数のフローを停止させるためにグローバルバックプレッシャを使うと停止させる必要のない他のフローまで停止させられてしまう。例えば、同一の入力ポートから入って異なる出力ポートへ出て行く二つのフローのうち片方のフローの

出力ポートだけが輻射しているときにグローバルバックプレッシャを実行すると、輻射箇所に関係のない本来影響を受けるべきでないもう片方のフローまで停止させられてしまう。品行の悪いフローの悪影響が通信装置を通過する全てのフローに及んでいる訳である。通信装置としては、停止すべきフローは停止させる一方、停止させるべきでないフロー（継続すべきフロー）はできる限りパケットを流すという公平さ（フェアネス）が必要である。つまり、従来技術ではフェアネスに欠けているところがある。

【0033】本発明では、フェアネスを実現できるようにするため、フローの停止が本来停止させなくて良いフローに広範囲に影響しないようにする手段を提供することを第3の目的としている。

【0034】

【課題を解決するための手段】第1の目的のために請求項1の通信方法においては、複数のフローを含むリンクによって互いに接続された通信装置間でフローを制御する通信方法において、予め、複数のフローをまとめたグループを定め、通信装置間で前記グループを通知しておき、下流の通信装置から上流の通信装置にフローの停止又は停止解除を通知するときは、前記グループによって通知することで、そのグループに含まれるフローに対して一括してフローの停止又は停止解除を通知することを特徴とする。

【0035】かかるように構成されているので、この通信方法は、類似した属性を備えたフローをグループとし、グループによって通知することで、それぞれのグループに含まれるフローを一括して停止又は停止解除する。

【0036】第1の目的のために請求項2の通信装置においては、複数のフローを含むリンクによって互いに接続される通信装置において、複数のフローをまとめたグループを示す下流側グループ参照手段と、下流の通信装置からグループによってフローの停止又は停止解除の通知を受ける フロー制御受信手段と、前記グループに含まれるフローに対してフローの停止又は停止解除を行うフロー制御実行手段とを有することを特徴とする。

【0037】かかるように構成されているので、この通信装置は、下流の通信装置からグループによってフローの停止又は停止解除の通知を受け、そのグループに含まれるフローを停止又は停止解除する。

【0038】第1の目的のために請求項3の通信装置においては、複数のフローを含むリンクによって互いに接続され、フローを構成するパケット又はセル（以下、パケットで代表する）を入力ポートから入力し一旦記憶領域に蓄えた後に出力ポートから出力することで通信を行う通信装置において、複数のフローをまとめたグループを示す上流側グループ参照手段と、グループに含まれるフローを構成するパケットに関して前記記憶領域に蓄わ

11

えられているパケット記憶量を求める統計手段と、前記パケット記憶量が停止閾値より多くなったときは上流の通信装置に前記グループによってフローの停止を通知し、前記パケット記憶量が再開閾値より少なくなったときは上流の通信装置に前記グループによってフローの停止解除を通知するフロー制御送信手段とを有することを特徴とする。

【0039】かかるように構成されているので、この通信装置は、グループに対するフローの停止又は停止解除を判断し、上流の通信装置に通知する。

【0040】第2の目的のために請求項4の通信装置は、複数のフローを含むリンクによって互いに接続され、フローを構成するパケット又はセル(以下、パケットで代表する)を入力ポートから入力し一旦記憶領域に蓄えた後に出力ポートから出力することで行う通信装置において、複数のフローをまとめたグループを示す上流側グループ参照手段であって、パケットの廃棄よりも遅配が望ましいグループ(以下、遅配グループ)と、パケットの遅配よりも廃棄が望ましいグループ(以下、廃棄グループ)とを示すものと、グループに含まれるフローを構成するパケットに関して前記記憶領域に蓄えられているパケット記憶量を求める統計手段であって、前記遅配グループと前記廃棄グループそれぞれに関して前記パケット記憶量を求めるものと、前記遅配グループのパケット記憶量が停止閾値より多くなったときは上流の通信装置に前記グループによってフローの停止を通知し、前記パケット記憶量が再開閾値より少なくなったときは上流の通信装置に前記グループによってフローの停止解除を通知するフロー制御送信手段と、前記廃棄グループのパケット記憶量が廃棄閾値より多くなったときは廃棄グループに含まれるフローのパケットを廃棄するパケット廃棄手段とを有することを特徴とする。

【0041】かかるように構成されているので、この通信装置は、パケット記憶量が所定の閾値より多くなったとき、遅配グループに対してはフローの停止を上流の通信装置に求め、廃棄グループに対してはパケットの廃棄を行う。すなわち、遅配グループと廃棄グループに対して、フローの属性に応じた異なる制御を行うことができる。

【0042】請求項5の通信装置は、廃棄されやすさの異なるパケットが前記廃棄グループのフローに含まれており、前記パケット廃棄手段は、パケットに含まれている情報を基に予め定められたきまりに従って廃棄されやすさを決定し、前記パケット記憶量に応じて廃棄されやすいパケットを優先的に廃棄することを特徴とする請求項4に記載の通信装置である。

【0043】かかるように構成されているので、この通信装置は、廃棄されやすいパケットを優先的に廃棄する。従って、廃棄されにくいパケットを、廃棄されやすいパケットよりも、優先的に伝送することができる。

12

【0044】請求項6の通信装置は、前記廃棄閾値が、前記記憶領域の空き容量に応じて変化することを特徴とする請求項4又は5に記載の通信装置である。

【0045】かかるように構成されているので、この通信装置は、記憶領域の空き容量に応じて廃棄閾値を変化させる。そのため、記憶領域の無駄を減らしつつ、状況に応じた制御が実現しうる。

【0046】請求項7の通信装置は、同一の出力ポートを通る複数のフローをまとめたグループがあり、各グループには、パケット記憶量の上限と、前記パケット記憶量の上限よりも小さい停止閾値及び再開閾値を備えることを特徴とする請求項3から6に記載の通信装置である。

【0047】かかるように構成されているので、この通信装置は、同一の出力ポートを通る複数のフローをまとめたグループでフローの停止と停止解除を制御できる。従って、ある出力ポートに対するグループのフローを停止する制御が、それら、別の出力ポートを通るフローを停止させる制御にはならない。

【0048】請求項8の通信装置は、グループに対するパケット記憶量が停止閾値より多くなった後、上流の通信装置にフローの停止を通知して、実際にフローが停止するまでの遅延時間を停止遅延時間とし、前記上流の通信装置が接続されている入力ポートから流入されるパケットの単位時間当たりの伝送量を伝送帯域とし、入力ポート毎に決まる停止遅延時間と伝送帯域の積に相当する容量を停止遅延相当分容量とし、グループに含まれるフローに関連する全入力ポートの停止遅延相当分容量の和に相当する容量をグループ停止遅延相当容量とすると、パケット記憶量の上限と停止閾値の差がグループ停止遅延相当容量以上であることを特徴とする請求項7に記載の通信装置である。

【0049】かかるように構成されているので、この通信装置は、フローの停止を通知してから、実際にフローが停止するまでに送られてくるパケットを、すべて記憶領域に記憶することができる。

【0050】

【発明の実施の形態】(第1の実施例)本発明による通信装置は、複数のフローに対して一括してフローの停止・停止解除を指定するバックプレッシャ(以下、グループバックプレッシャと呼ぶ)を新たに用いる。従来技術で述べた個々のフローに対する選択的バックプレッシャに加えてグループバックプレッシャを用いることで、個々のフローの制御に加えて特定の属性を共通に備える複数のフローを集約的に制御する。このことにより、フロー間のフェアネスを保証しつつ、類似した属性を有するフロー・グループ各々に対して最適な制御を提供することができる。

【0051】本発明を実施するためのグループバックプレッシャは、図4に示した従来技術と同様の構成の通信

13

装置に適用することができる。ここでは、バックプレッシャの動作を説明する便宜のため、下流の通信装置につながる一つの出力ポートに着目して通信装置を図1に示す。なお、従来技術でも述べたように、本発明の通信装置は、パケットを送送する場合でもセトルを送送する場合でも適用可能であり、セトルの場合を含めてパケットで代表して説明する。また、本発明の通信装置では、フローの最低保証帯域を率に限定しておらず、フローにはベスト・エフォート・サービスを適用しないもの(例:帯域予約型フロー)も含んでいて良いものとする。以下では、特にベスト・エフォート・サービスを利用するフローに関する制御について述べる。

【0052】図1において、35はlink-by-linkのバックプレッシャによるフロー制御を行う通信装置である。47、48、49は、それぞれ複数のフローを含むリンクであり、接続されている装置と両向きに双方向の通信を行う全二重(full-duplex)動作を行う。41、42、43は入力ポートであり、44、45、46は出力ポートである。

【0053】図1はリンク49に接続されている通信装置36の出力ポート46の視点で描かれている。すなわち、出力ポート46から出力されるパケットは、入力ポート41と42から入力されるものとする。そして、リンク47、48の接続先の通信装置は上流側の通信装置、リンク49の接続先の通信装置は下流側の通信装置と呼ぶことにする。下流側の通信装置から本通信装置36へのフロー制御パケットは入力ポート43から入力される。本通信装置36から上流側の通信装置へのフロー制御パケットは出力ポート44、45から出力する。

【0054】通信装置35のリンク47～49に別の通信装置36～38をつなない状態を図2に示す。説明の便宜上、図2に示す各通信装置35～38の内部は、後述する上流側グループ参照テーブル81、82、85、87と、下流側グループ参照テーブル83、84、86、88のみ示している。

【0055】図1においては、説明の便宜上、上記以外のパケットが通過する経路は省略して示している。例えば、47、48以外のリンクから入力されるパケットや49以外のリンクへ出力されるパケットの通過するリンクや出力ポートは省略されている。また、リンク49の接続先の通信装置から本通信装置36を経てリンク47の接続先の通信装置に送られるパケットに着目すると、上記の上流・下流の関係は逆になるが、かかる場合のバックの経路は省略されている。言うまでもなく、これら図示を省略した経路におけるフロー制御についても、それぞれ本発明は実施できる。なお、入力ポート43から入力されるB-Pパケット受信部55に伝えられるフロー制御パケットは、図5に示すようにスイッチコア51を経由するが、分かりやすくするため、図1においては途中を省略して示している。同様に、B-Pパケット送

14

信部58から出力ポート44又は45に伝えられるフロー制御パケットは、図5に示すようにスイッチコア51を経由するが、分かりやすくするために、図1においては途中を省略して示している。

【0056】図1に示す51はスイッチコアであり、複数の入力ポート(うち、図1では入力ポート71と72のみを示す)と複数の出力ポート(うち、図1では出力ポート75のみを示す)を備えている。スイッチコア51は、入力されたパケットをバッファ52へ書き込むとともに、パケットが属するフローを識別し、フロー識別子とパケット記述子をキュー管理部53に通知する。また、スイッチコア51は、キュー管理部53からフロー識別子とパケット記述子を受け取ると、そのパケットをバッファ52から取り出し、フロー識別子に対応する出力ポートからパケットを出力する。本実施例では、フロー識別子は、フローを互いに区別して表す数値(フロー番号)で実施することとする。

【0057】52は全パケットを一時的に記憶するバッファである。パケットをバッファ52のどこに記憶したかは、パケット記述子に示されている。パケット記述子の大きさは、パケットよりも小さい(バイト数が少ない)ので、パケットそのものをキューに蓄えて時分処理を行う場合よりも、キューの記憶容量が節約でき、キューの操作が容易になる。

【0058】53はキュー管理部である。キュー管理部53は専用のキュー管理用メモリ60を使用してフロー毎のキューを管理する。スイッチコア51にパケットが入力されて、スイッチコア51からフロー識別子とパケット記述子をキュー管理部53が受け取ると、そのフロー識別子で示されるキューにそのパケット記述子をキューイングする。このとき、パケット記述子が空のキューに対してキューイングされたのであれば、キュー管理部53はスケジューラ54にフローが活性化したことを通知する。

【0059】また、キュー管理部53はスケジューラ54からフロー識別子を受け取ると、そのフローに対応するキューの先頭に位置するパケット記述子を取り出す。このとき、キューが空になれば、キュー管理部53はスケジューラ54にフローが非活性化になったことを通知する。また、スイッチコア51にフロー識別子とパケット記述子を通知することでパケットの出力を指示する。

【0060】54はスケジューラである。スケジューラ54はフロー識別子と出力ポートの対応、出力ポートの有する伝送帯域を各フローにどれだけ割り当てるか、を管理している。スケジューラ54は、キュー管理部53から活性化したフローの通知を受けると、そのフローをサービスの対象に加える。また、スケジューラ54がキュー管理部53から非活性化したフローの通知を受けると、そのフローをサービスの対象から除外する。スケジューラ54は、それぞれの出力ポートに属するサービス

15

対象のフローに関して、既定の伝送帯域割り当てポリシーに従って、パケットを出力するフローを逐次選択し、キュー管理部53に通知する。この際、後述するようにバックプレッシャによりフローの停止・停止解除の状態を加味する。

【0061】81は上流側グループ参照テーブルであり、複数のフローをまとめたグループを示す。グループは、例えば、同じ入力ポートを通るフローをまとめたもの、同じ出力ポートを通るフローをまとめたもの、パケットの廃棄よりも選配が望ましいフローをまとめたもの（選配グループ）、パケットの選配よりも廃棄が望ましいフロー（廃棄グループ）をまとめたもの、など種々の観点から設定できる。本発明ではグループバックプレッシャを実施するので、上流側グループ参照テーブル81の内容は、リンク47、48につながる上流の通信装置と一貫させる必要がある。そのため、上流側グループ参照テーブル81の内容（グループとそれに含まれるフローの対応）を、図2に示す通信装置36、37に通知してそれぞれの下流側グループ参照テーブル84、86に加えておけば良い。

【0062】図3(a)に通信装置35の上流側グループ参照テーブル81の内容を例示する。グループとして、選配グループ、出力ポート1グループ（第1の出力ポートを通るフローをまとめたもの）、出力ポート2グループ（第2の出力ポートを通るフローをまとめたもの）の3つが示されている。通信装置間でグループを示すために、グループには番号（グループ番号）が1、2、3、…と付されている。通信装置35を通るフローにはフロー番号が1、2、…と付されている。各フローに付した、1か0によってグループに属する、属さないの区別が示されている。例えば、選配グループには、2と3のフロー番号のフローが該当する。

【0063】図3(b)に、図2に示す通信装置36の下流側グループ参照テーブル84の内容を例示する。通信装置36の出力ポート1には通信装置35が接続されており、図には示していないが、出力ポート2にはさらに別の通信装置が接続されている。図3(b)は、通信装置36のフロー番号4、5、6、7、8が、通信装置35におけるフロー番号1、2、3、7、8にそれぞれ対応する場合を示している。従って、図3(a)のグループ番号1、2、3におけるフロー番号1、2、3、7、8の値（0か1）が、図3(b)のグループ番号1、2、3におけるフロー番号4、5、6、7、8に、それぞれ写されている。なお、この例では、通信装置35におけるフロー番号と通信装置36におけるフロー番号の対応を示す別のテーブルをどちらかの装置内に設けておく。

【0064】57は統計収集部である。統計収集部57は、キュー管理部53から通知を受けて、フロー毎キュー61〜64のキュー長や、各フローに割り当てた伝送

16

帯域の実績、活性化しているキューの個数、バッファ52の使用量といった情報を収集する。そして、後述するような方針に従って、バックプレッシャの実行を判断する。すなわち、個々のフロー、グループ単位でのフロー、あるいは、全フローに対する、停止・停止解除を判断する。停止・停止解除が必要と判断した場合は、それぞれ、個々のフロー番号、グループ番号、あるいは、全フローである旨を、B/Pパケット送信部58へ通知する。

【0065】バックプレッシャを実行する方針は、選択的バックプレッシャとグローバルバックプレッシャについては、従来技術と同様に実施できる。例えば、キュー61〜64に対して選択的バックプレッシャの停止閾値と再開閾値を設けておく。キュー長が停止閾値より大きくなればフローを停止する。逆に、キュー長が再開閾値より小さくなればフローを停止解除する。また、バッファ52に対してグローバルバックプレッシャの停止閾値と再開閾値を設けておく。バッファ52のバケット記憶量が停止閾値より大きくなれば全フローに対する停止を行う。逆に、バケット記憶量が再開閾値より小さくなれば全フローに対する停止解除を行う。

【0066】グループバックプレッシャを実行する方針は、バッファ52に蓄えられているバケットの数をグループ毎に集計し、それぞれのグループ毎に設けた停止閾値と再開閾値と比較することで実施できる。すなわち、まず、グループに含まれるフローに関してバッファ52に蓄えられているバケットの量（バケット記憶量）を求める。そして、このバケット記憶量が停止閾値より大きくなればそのグループに対する停止を行う。逆に、このバケット記憶量が再開閾値より大きくなればそのグループに対する停止解除を行う。

【0067】58はB/P（バックプレッシャ）パケット送信部である。B/Pパケット送信部58は統計収集部57からの通知を受けて、フローの停止あるいは停止解除を通知するバックプレッシャパケットを作成する。そして、バックプレッシャを実行すべき上流の装置に対して、出力ポートを経て、そのバックプレッシャパケットを送出する。

【0068】55はB/P（バックプレッシャ）パケット受信部である。B/Pパケット受信部55は下流の装置から送出されたバックプレッシャパケットを受信する。そして、そのバックプレッシャパケットが正当であれば、B/P実行部56へ引き渡す。

【0069】58はB/P（バックプレッシャ）実行部である。B/P実行部56は、B/P受信部55から受け取ったバックプレッシャパケットを解釈して、スケジューラ54に対してフローの停止、停止解除を指示する。グループバックプレッシャにより、グループに対する停止、停止解除が通知されている場合には、下流側グループ参照テーブル82を参照することで、そのグループに含ま

50

17

れる個々のフロー番号を得る。そして、それらのフローを停止、停止解除の対象とする。グループバックプレッシャによる停止処理では、従来技術のようにスケジューラを停止させる必要はない。言うまでもなく、従来技術と同様に、選択的バックプレッシャの場合はそのバックプレッシャパケットで示されるフローを停止、停止解除の対象とする。また、グローバルバックプレッシャの場合は全フローを停止、停止解除の対象とする。

【0070】82は下流側グループ参照テーブルである。下流側グループ参照テーブル82はグループ番号とそのグループに含まれるフローを示している。リンク49につながる下流側の通信装置とグループの定義が一貫している必要がある。そのためには、例えば、図2に示す通信装置38の有する上流側グループ参照テーブル87の内容を受け取って、その中から、リンク49を通るフローに関するグループを、通信装置35の下流側グループ参照テーブル82に加えておきたい。

【0071】本実施例においては、スケジューラ54が、キュー管理部53に対してパケットの取り出しを示す(すなわち、フローをサービスの対象とする)か否かは、次に示す4つの条件の論理積に従う。

- (a) フローが活性化されている(キューが空でない)。
- (b) 選択的バックプレッシャによるフローの停止を受けていない。
- (c) グループバックプレッシャによるフローの停止を受けていない。
- (d) グローバルバックプレッシャによるフローの停止を受けていない。

【0072】すなわち、(a)かつ(b)かつ(c)かつ(d)が成り立つようになったら、そのフローはスケジューラ54のサービスの対象に追加される。そして、スケジューラ54はそのフローに既定のポリシー(例えば、DDR、WRR)に従ってパケットの出力を指示する。

【0073】逆に、(a)かつ(b)かつ(c)かつ(d)が成り立たなくなったなら、そのフローはスケジューラ54のサービスの対象から除外される。すなわち、フローが非活性化された、又は、選択的バックプレッシャによるフローの停止を受けた、又は、グループバックプレッシャによるフローの停止を受けた、又は、グローバルバックプレッシャによるフローの停止を受けた、ならば、パケットの出力を指示しない。

【0074】なお、グローバルバックプレッシャは、従来技術と同様にスケジューラを停止させることによって実行できる。あるいは、全部のフローをスケジューラ54のサービスの対象から除外することも実行できる。

【0075】また、グローバルバックプレッシャを用いない通信装置を実施する場合は、条件(b)は無い。この場合は、(a)かつ(b)かつ(c)が成り立てばそのフローをスケジューラのサービスの対象に加え、(a)かつ(b)かつ(c)が成り立たなくなればそのフローをスケジューラの

18

サービスの対象から除外すれば良い。

【0076】以上の説明では、グループバックプレッシャによるフロー制御を実行するときは、バックプレッシャパケットにおいてグループ番号を指定することで、そのグループに属する複数のフローに対してフローの停止・停止解除を一括して指示している。

【0077】グループを指定する方法は、必ずしもグループ番号に限る必要はない。例えば、フローに属性をつけておき、属性とその属性を有するフローを下流側グループ参照テーブルに登録しておく。すると、バックプレッシャパケットではフローの属性を指定することで複数のフローを一括して指定できるようになる。

【0078】あるいは、フロー番号の値によって、フローの属性が表されるようにフロー番号の割り振り方を工夫すれば、上流側グループ参照テーブル81と下流側グループ参照テーブル82に記載する内容を簡略化することもできる。例えば、フロー番号を5桁の16進数で表すことにし、最上位の1桁目の値が1なら選配グループ、最上位から数えて第2桁目の値が1なら出力ポート1グループ、最上位から数えて第2桁目の値が2なら出力ポート2グループ、最上位3桁は個々のフロー毎の通し番号、と定めておく。すると、グループとして「x1xx x」を指定すれば(xは、don't careの意味)、グループバックプレッシャの対象は出力ポート1グループになる。かかる場合は、フロー番号の割り振り方の規則を、上流側グループ参照テーブル81と下流側グループ参照テーブル82に記載しておきたい。

【0079】グループ番号やフローの属性を使うのは、明示的なグループで複数のフローを指定する方法であるが、明示的にグループを指定せずに複数のフローを指定する方法もある。例えば、グループバックプレッシャを指示するバックプレッシャパケットにおいて、各フローを各ビット位置に対応させたデータを送るとする。すると、このデータの0、1のビットパターンで、複数のフローに対して停止・停止解除を一度に指示することができる。かかる場合は、図1に示す下流側グループ参照テーブル82には、各ビット位置とフローとの対応関係を設定しておきたい。

【0080】以上の説明では、上流側グループ参照テーブルをまず定め、その内容を上流側の通信装置にある下流側グループ参照テーブルに加えるように説明した。グループはリンクでつながれた通信装置間で一貫していれば良いので、これを逆に、下流側グループ参照テーブルをまず定め、その内容を下流側の通信装置にある上流側グループ参照テーブルに加えるようにしても良い。また、通信装置の内部でグループを定めるのではなく、通信装置の外部に設けた管理機構(例えば、ネットワークを管理するネットワーク管理装置)がグループを定めて、各通信装置の上流側グループ参照テーブルと下流側グループ参照テーブルに設定するのでも良い。

50

19

【0081】なお、通信装置35内では、上流側グループ参照テーブル81と下流側グループ参照テーブル82を別々に用意するのではなく、両者の内容を統合したグループ参照テーブルを設けておき、これをB/P実行部56と統計収集部57が使用するのでも良い。

【0082】以上の説明から判るように、請求項1の発明は、図2に示す通信装置35において上流側グループ参照テーブル81を定め、その内容を上流側の通信装置36に通知しておき、グループバックプレッシャを行うことで実施される。あるいは、通信装置35において下流側グループ参照テーブル82を定め、その内容を下流側の通信装置38に通知しておき、グループバックプレッシャを行うことでも実施される。あるいは、ネットワーク管理装置がグループを定め、通信装置間の通信によってそのグループを通知し、グループバックプレッシャを行うことでも実施される。言うまでもなく、上流側グループ参照テーブルと下流側グループ参照テーブルを統合したグループ参照テーブルを用いる場合も、同様に実施される。

【0083】請求項2の発明において、下流側グループ参照手段は、図1に示す下流側グループ参照テーブル82で実施される。フロー制御受信手段は、B/Pパケット受信部55で実施される。フロー制御実行手段は、B/P実行部56で実施される。

【0084】請求項3の発明において、上流側グループ参照手段は、図1に示す上流側グループ参照テーブル81で実施される。統計手段は、統計収集部57で実施される。フロー制御送信手段は、B/Pパケット送信部58で実施される。

【0085】(第2の実施例) 請求項4の発明の実施例を次に説明する。図1に示す上流側グループ参照テーブル81によって上流側グループ参照手段が実施され、統計収集部57によって統計手段が実施され、B/Pパケット送信部53によってフロー制御送信手段が実施されるのは、第1の実施例と同様である。第1の実施例に対して、さらに、次の工夫を施すことによって特定の効果を得ることができる。

【0086】まず、ベスト・エフォート・サービスを利用するフローを、パケットの廃棄よりも遅配が望ましいグループ(遅配グループ)と、パケットの遅配よりも廃棄が望ましいグループ(廃棄グループ)とに分類する。これらのグループを、上流側グループ参照テーブル81に登録しておく。そして、遅配グループに対しては、その遅配グループに含まれるフローに関してバッファ52に蓄えられているパケットの量(パケット記憶量)を統計収集部57において求める。もし、このパケット記憶量が停止閾値より多くなったと統計収集部57が判断したなら、遅配グループのフローの停止を求めるバックプレッシャパケットをB/Pパケット送信部58が作成して上流の通信装置に送る。また、パケット記憶量が再開閾

20

値より少なくなった統計収集部57が判断したなら、遅配グループのフローの停止解除を求めるバックプレッシャパケットをB/Pパケット送信部58が作成して上流の通信装置に送る。

【0087】一方、廃棄グループに対しては、廃棄閾値を設けておき、パケット記憶量が廃棄閾値より多くなったと統計収集部57が判断した場合には、パケットの廃棄を行う。統計収集部57の判断はキュー管理部53を経てスイッチコア51が受け取り、スイッチコア51がパケットの廃棄を実行する。すなわち、スイッチコア51はパケット廃棄手段を兼ねている。具体的なパケットの廃棄は、例えば、後から到着したパケットに関して、バッファ52への書きこみを行わないことと、パケット記述子及びフロー識別子をキュー管理部53に通知しないことでも実施する。

【0088】個々のフローに対しては、従来技術と同様に、フロー毎キュー61~64に蓄えられているキュー記憶量に基づく選択的バックプレッシャを行う。なお、グローバルバックプレッシャは行わない。

【0089】このように実施することで、遅延グループに対してグループバックプレッシャを実行しても、廃棄グループに含まれるフローが停止させられることがなくなる。すなわち、リアルタイム性の高いフローは廃棄グループに含めておけば、従来技術によるグローバルバックプレッシャで生じたような、他のフローの影響を受けてフローが中断させられる事態を、防ぐことができる。

【0090】なお、廃棄グループによって廃棄閾値を設けて、一定以上のパケットがバッファ52に蓄積されないようにする理由は、遅延グループのフローを停止してそのパケット記憶量が減った際に、本来遅延グループのために確保しておくべきバッファ52の容量を廃棄グループが占有しないためでもある。

【0091】(第3の実施例) ここで、パケットの遅配よりも廃棄が望ましい用途(アプリケーション)を例示しておく。この用途は、例えば、階層的に符号化された画像をリアルタイムにマルチキャスト配信するアプリケーションが該当する。階層的に符号化された画像配信の概要は次の通りである。

【0092】まず、低解像度の画像と共に、段階的に解像度が増す高解像度画像を準備する。高解像度の画像は伝送できない場合は諦めても良いが、少なくとも低解像度の画像は伝送したい。そこで、アプリケーションは低解像度画像を高い優先度を付したパケットとし、段階的に用意された高解像度画像との差分情報をそれぞれ段階的に低くなる優先度を付したパケットとしてネットワークへ出力する。すなわち、優先度が高ければそのパケットは廃棄されるが、優先度が低ければそのパケットは廃棄されても支障はより少ない。

【0093】請求項5の発明の実施例を次に説明する。本実施例の通信装置においては、上記パケットを含むフ

50

ローを廃棄グループに分類し、パケット記憶量が廃棄閾値を越えたときは、優先度の低いパケットをまず廃棄する。優先度の低いパケットを廃棄してもなお、図1に示すバッファ52の余裕が少ない場合は、より優先度の高いパケットも廃棄する。すなわち、パケット記憶量と廃棄閾値の差に対応して、どの優先度のパケットまで廃棄するかを定めておく。バッファ52の余裕は、統計収集部57で求め、その余裕はキュー管理部53を経てスイッチコア51に伝え、スイッチコア51が廃棄処理を行う。

【0094】結果として、伝送帯域に余裕があるリンクに接続される通信装置では高解像度の画像が伝送され、伝送帯域に余裕が無いリンクに接続される通信装置では低解像度の画像が伝送される。あるいは、同一のリンクを使う場合であっても、伝送帯域に余裕のある時間帯では高解像度の画像が伝送され、伝送帯域に余裕が無い時間帯では低解像度の画像が伝送される。

【0095】このようなアプリケーションはパケットの廃棄を前提として設計されている。配信される画像がリアルタイム性の高いものであれば、パケットが選配される
20 ときの優先順位もたまたまいまいか、経路の途中の帯域を消費することによって他のフローに対して迷惑をかける原因となる。本実施例においては、かかるパケットの選配が生じないので、他のフローに対する支障を未然に防ぐことができる。

【0096】なお、パケットの廃棄のされやすさをセケル（セル）内に明示した実施例には、ATMのCLP（Cell loss priority）が該当する。また、パケットの廃棄されやすさがパケット内に明示されてなくても、プロトコルの種類によって廃棄されやすさを定めることで本発明は実施できる。例えば、IP（Internet Protocol）パケットは、TCP（Transmission Control Protocol）やUDP（User Datagram Protocol）のデータグラムを含んでいる。TCPはパケット廃棄に着目した伝送量の制御を行っており、パケットが欠落したときは、パケットを再送することも行っている。これに対して、UDPは伝送量の制御や、パケット損失時のパケット再送を行っておらず、リアルタイム通信によく使われる。そこで、UDPによるパケットよりも、TCPによるパケットの方が、パケットは廃棄されやすくと定めておく。すると、本発明を実施するには、スイッチコア51において、パケットの種類を区別して、UDPよりもまずTCP
30 のパケットを廃棄すればよい。

【0097】（第4の実施例）請求項6の発明の実施例を次に説明する。第2の実施例における廃棄グループの廃棄閾値を、バッファ52の空き容量が多い場合には高くし、空き容量が少ない場合には低くする。例えば、バッファ52の空き容量をすべて廃棄グループに割り当てる。あるいは、バッファ52の空き容量を、廃棄グループを含めた幾つかのグループに割り振るのも良い。す
50

ると、空き容量が多い場合には、パケットの廃棄が起りにくくなる。

【0098】通信装置に送られてくるフローの数やパケットの量は、いつも一定しているのではなく、時間によって増えたり減ったりするので、記憶領域の空きは増えたり減ったりする。一般に、一つのフロー当たりの記憶容量を減らせば少ない記憶領域でも多くのフローが扱え、一つのフロー当たりの記憶容量を増やせばパケットの廃棄が生じにくくなる。従って、記憶領域の空きが大きいときは、廃棄閾値を上げてより多くの記憶容量を使うことでパケットの廃棄を生じにくくできる。記憶領域の空きが少なくなったときは、廃棄閾値を下げるのでフローの数を減らす必要はない。

【0099】このように、バッファ52の空き容量は時間とともに変化するが、本実施例においては、空き容量が多い場合にはその空き容量を有効に使うことができる。その結果、平均的に見れば、廃棄閾値を一定値に固定する場合（余裕を見込んで低めに設定せざるをえない）に比べると、廃棄閾値は高くなるので、パケットの廃棄が起りにくくなる。

【0100】（第5の実施例）請求項7の発明の実施例を次に説明する。本実施例では、第1から4に示した実施例において、まず、ベスト・エフォート・サービスを利用するフローを出力ポート別のグループに分け、上流側グループ参照テーブル81に登録しておく。各グループには、バッファ52内に占有できるパケット記憶量の上限を設定し、さらに、その上限値よりも小さい停止閾値と再開閾値を統計収集部57に設けておく。グループバルバックプレッシャは実施しないが、個々のフローに対する選択的バックプレッシャは実施する。すなわち、B
Pパケット送信部58は、選択的バックプレッシャグループバックプレッシャのバックプレッシャパケットを必要に応じて送信する。

【0101】以上のよう構成することで、輻輳している出力ポートに関連するフローのみをグループバックプレッシャにより停止させることができる。出力ポート毎に使用可能なバッファ量の上限を設ける理由は、輻輳しているポートに関連するグループのパケットがバッファを占有してしまうことを防ぐためである。

【0102】（第6の実施例）請求項8の発明の実施例を次に説明する。第5の実施例に示すグループ毎のパケット記憶量の上限と停止閾値の差を、次のように選ぶことによってグループバックプレッシャによるフローの停止時のパケット廃棄を防ぐことができる。

【0103】まず、グループ毎に確保されるパケット記憶量の上限に相当する容量をバッファ52内で排他的に確保しておく。次に、停止閾値を次のように選ぶ。グループ上流の通信装置にそのグループに属するフローの停止を要求するグループバックプレッシャを通知してから、実際にフローが停止するまでの遅延時間を停止遅延

時間とする。一方で、その入力ポートから流入されるパケットの単位時間当たりの伝送量を伝送帯域とする。この停止遅延時間と伝送帯域は、入力ポート毎に決まる値である。それから、停止遅延時間と伝送帯域の積に相当する容量を入力ポートの停止遅延相当分容量とする。グループに含まれるフローが通過している全入力ポートの停止遅延相当分容量の和に相当する容量をグループ停止遅延相当容量とする。このグループ停止遅延相当容量がパケット記憶量の上限と停止閾値の差となるように停止閾値を設定する。この停止閾値を、統計収集部57に設定しておく。もちろん、安全のために多少の余裕を見込んで、パケット記憶量の上限と停止閾値の差がグループ停止遅延相当容量よりも多くなるように、停止閾値を設定しても良い。

【0104】なお、パケット記憶量の上限と停止閾値をバッファ52の空き容量に応じて動的に変化させる実施例も可能である。例えば、バッファ52の空き容量をそれぞれのグループに割り振る。すると、バッファ52の空き容量が多い場合には、パケット記憶量の上限と停止閾値は高くなり、空き容量が少ない場合には低くなる。このように、動的に変化させることによって、バッファ52の空き容量がある場合には、パケットの廃棄を起こりにくすることができる。

【0105】

【発明の効果】請求項1から3に記載の発明においては、複数のフローからなるグループを対象としてフローの制御が行われる。そのため、個々のフローを指定する場合に比べると、同じグループに属するフローの停止又は停止解除の通知を、より簡潔なバックプレッシャパケットで実施できる。従って、バックプレッシャパケットの伝送量及び伝送時間を低減することができる。

【0106】請求項4に記載の発明においては、選配グループに属するフローはパケットの廃棄ではなくフローの停止が行われる。一方、廃棄グループに属するフローは、フローの停止ではなくパケットの廃棄が行われ、フローが完全に中断されることはない。すなわち、それぞれのフローの属性に応じた取り扱いが実現できる。

【0107】請求項5に記載の発明によれば、さらに、廃棄されやすさに応じたパケットの廃棄が行われる。すなわち、廃棄が生じる場合であっても、廃棄されるべきでないパケットを優先的に伝送することができる。

【0108】請求項6に記載の発明においては、記憶領域の空き容量に応じて廃棄閾値を変える。このことで、記憶領域の空き容量を有効に使って、パケットの廃棄が起こ

りにくくすることができる。

【0109】請求項7に記載の発明においては、出力ポートに対応したグループを単位としてフローの停止及び停止解除を実行することができる。従って、輻輳している出力ポートに関連するフローのみを停止させることができる。すなわち、輻輳箇所に関係のないフローが不当に停止させられることが防げる。

【0110】請求項8に記載の発明においては、フローの停止を通知してから、実際にフローが停止するまでに送られてくるパケットを記憶するだけの容量が記憶領域に確保されている。従って、フローの停止によってパケットの廃棄が生じることがない。

【図面の簡単な説明】

【図1】本発明による通信装置の構成を示す図である。

【図2】本実施例による通信装置の接続を示す図である。

【図3】図3(a)と図3(b)は、それぞれ、上流側グループ参照テーブルと下流側グループ参照テーブルの内容例を示す図である。

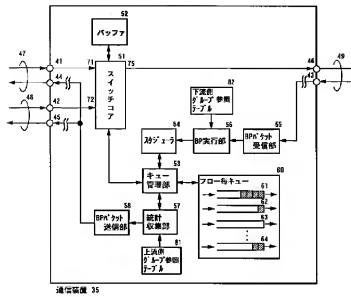
【図4】ネットワークの構成例を示す図である。

【図5】従来技術による通信装置の構成を示す図である。

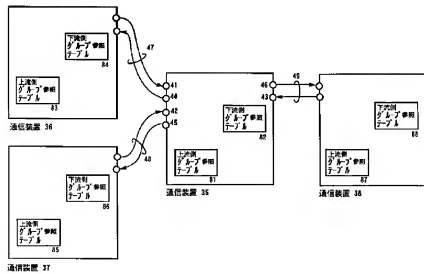
【符号の説明】

30、31、32、33、34：コンピュータ
35、36、37、38：通信装置
41、42、43：入力ポート
44、45、46：出力ポート
47、48、49：リンク
51：スイッチコア
52：バッファ
53：キュー管理部
54：スケジューラ
55：B/Pパケット受信部
56：B/P実行部
57：統計収集部
58：B/Pパケット送信部
60：キュー管理用メモリ
61、62、63、64：キュー
71、72、73、74：スイッチコア51の入力ポート
75、76、77、78：スイッチコア51の出力ポート
81、83、85、87：上流側グループ参照テーブル
82、84、86、88：下流側グループ参照テーブル

【図 1】



【図 2】



【図3】

(a) 上流側グループ参照テーブル

グループ名	グループ番号	フロー番号							
		1	2	3	4	5	6	7	8
通配グループ	1	0	1	1	0	0	0	0	...
出力ポート1グループ	2	1	0	1	1	1	0	0	...
出力ポート2グループ	3	0	1	0	0	0	1	1	...
.
.
.

(b) 下流側グループ参照テーブル

下流側	グループ番号	フロー番号							
		1	2	3	4	5	6	7	8
出力ポート1	1	0	0	0	0	1	1	0	...
	2	0	0	0	1	0	1	0	...
	3	0	0	0	0	1	0	1	...

出力ポート2	1	1	0	1	0	0	0	0	...
	2	0	1	0	0	0	0	0	...
	3	1	0	1	0	0	0	0	...

.
.
.

【図4】

